REVIEW ARTICLE

Design and Analysis of Observational Studies

[1]KP Suresh, [2]MR Gajendragad, [3]H Rahman

ABSTRACT

Appropriate study design forms the basis of any successful organized research. Irrespective of the underlying rationale of the research question, inappropriate design of an experiment or execution of a study may not give a valid result. As defined, the research should have the following characteristics: it should be undertaken within the framework of a set of philosophies, should have an unbiased objective and a design comprising of methods and techniques that have been tested for their validity and reliability. These principles of research clearly emphasize the need for an appropriate study design. The current review focuses on the planning and analysis of an observational study to obtain significant and accurate findings.

Keywords: Experimental design, Research methodology, Analysis, Observational studies.

INTRODUCTION

Observational studies can be broadly categorized into three types: cross-sectional, case-control or prospective. Each of these three study types have their own merits and demerits and the present review discusses the benefits and drawbacks of each type in detail.

In cross-sectional studies (CSS), the measurements are taken at the same time and does not involve manipulation of variables. Such studies, are mainly employed for evaluating the prevalence of a variable of interest in a population. It allows researchers to measure multiple measurements at once, and is good for descriptive analysis and generating newer hypothesis. The prevalence obtained is important for the public health assessment of the disease burden in a specified population and in planning and allocating health resources. For example, in a CSS we assess burden of high

[1]Scientist, [2]Principal Scientist, [3]Director

[1-3]ICAR-National Institute of Veterinary Epidemiology and Disease Informatics, Bengaluru, Karnataka, India

Corresponding Author: KP Suresh, Scientist, ICAR-National Institute of Veterinary Epidemiology and Disease Informatics Bengaluru, Karnataka, India, e-mail: sureshkp97@gmail.com

blood pressure and presence of heart disease at same point of time, and look for possible association between blood pressure and heart disease and if found, it would be difficult to say which occurred first. Did the heart disease result in high blood pressure or blood pressure caused the heart disease or heart disease and high blood pressure both together caused some other disease(s). The CSS can be conducted using several mode of data collection *viz* telephone interview, face-to-face interview, mailed questionnaire, electronic mail, web data collection, any other self-administered questionnaire. Though the design appears very simple, enlisting participants who are very similar in one specific variable can be difficult. The groups can be affected by cohort differences that arise from particular experiences of unique group of patients or participants. For example, group of participants born in the same time period share important historical experiences, while other group born in specific geographic locations share experiences limited to their physical location. Other disadvantage of this design is the difficulty in determine whether outcome followed the exposure in time or exposure resulted from the outcome. The CSS are primarily designed to measure prevalence rather than incidence and not suitable for measuring the rare disease or disease with short duration. The associations identified by the design are difficult to interpret and is influenced by low response and misclassification due to recall bias.

Case-control studies (CCS) are often referred to as retrospective study because the data on factor of interest (*viz* the cause/determinant of the event) is collected retrospectively. They are useful for rare disease or conditions or when disease takes long time to become manifest. In a CCS, one group will be the subjects having disease or condition and the another group will be those without disease or condition and investigator goes backward into a past time to determine the factors or exposures associated with the disease development. For example, in a case-control study of smoking and lung cancer, the investigator starts the study with the group of patents having lung cancer and control group without lung cancer. Thorough examination of records and/or interviews are performed to determine the presence or absence of factors (smoking) that may influence the association of smoking with the occurrence of lung cancer. The main strengths of CCS are cost-effectiveness and possible completion in a short duration. CCS is useful to study the exposure or variables that do not change with

time and are commonly used for obtaining quick results and investigating outbreaks. Since, the case-control studies start with subjects with disease or condition, it is may be possible to enroll sufficient number of patients with rare disease. The disadvantages are that CCS is more expensive and less efficient. A limitation of CCS is the bias associated with selection of cases and controls from the same study population due to the difficulty in finding 'good match' controls. Bias may also arise due to the reliance on the recall of information that leads to a recall or bias. CCS is not suitable to evaluate the prognostic and etiological roles of exposure of interest. Selection of controls is more difficult in CCS as it is difficult to satisfy the matching criteria of risk factors exposures and confounders with that of cases. CCS is often used before conducting longitudinal studies.[1,2]

The prospective or cohort studies involve the assessment of an outcome in subjects subsequent to exposure or non-exposure to a treatment. The prospective or cohort studies are also known as longitudinal studies. The primary objective of cohort studies is to investigate the association between exposure and disease prospectively. For example, in a prospective study investigator starts with cohorts of nondiseased individuals who are exposed to smoking along with those unexposed and determine the risk of developing lung cancer in future. Here exposure is smoking and disease and outcome is lung cancer. The strengths of cohort studies are the calculation of disease incidence and evaluation of the effect of multiple exposures on the outcome, in a single study. Another major advantage of a cohort study is the ability to assess temporal relationships and document the events before the occurrence of outcome.

SELECTION OF SUBJECTS, CASES AND CONTROLS

One of the major issues associated with the design of an observational study is the selection of representative subjects or participants. Measurement of the entire population is not possible in any research especially those studies involving human subjects. Hence, it is easier and more practical to study selected samples from a predefined population. The target population depends on the research question to be answered and the sample represents a subset of the larger population to be studied. Different techniques for sampling are used in observational studies. The sampling frame is the population from which the sample is drawn which can even be the patients attending a specific hospital.

Probability sampling is the method wherein there is a finite chance for each participant being selected into study. Random sampling is one in which all the subjects within a predefined population have an equal probability of being included. Another technique is the systematic sampling in which the population is divided by the number of subjects required (k); then individuals are selected from a random starting point. In case, the sample required is 100 out of a population of 1000, every 10th individual will be selected with the first one being drawn from random number say 3. In stratified sampling, the population is divided into strata based on predefined population characteristics of interest, and participants are then randomly selected from each stratum. This ensures that each stratum is proportionally represented in the final sample. Cluster sampling is used when a using a population list is not feasible and the initial selection is performed using a number of a clusters or groups. The population is divided into homogeneous clusters in such a way that subjects within the clusters are as heterogeneous as possible. Cluster should be mutually exclusive and collectively exhaustive. In Multistage cluster sampling technique smaller clusters are randomly selected from previously selected larger clusters.

Nonprobability sampling methods do not ensure equal probability to include an individual in the population in the sample. Convenience sampling is one such method where participants are selected based on availability (e.g. hospital staff). If subjects who fit into a particular criterion are selected, the sampling is called quota sampling. Referral sampling (or snowballing) involves participants in the sample who have been referred by other participants. In a case control study, the selection of appropriate cases for evaluation and right control for comparison are important.

THE LOGIC OF OBSERVATIONAL STUDIES AND THE PROBLEM OF BIAS

Observational studies are preferred in circumstances where hypothesis cannot be evaluated by experiments, and are still possible to make argument for cause and effect though the usual procedures of statistics when laws of probability conferred by randomization cannot be applied. In the absence of random sampling and assignment of treatments or intervention, researcher must be careful in interpreting the results of these studies, recognizing the potential for error. Thus, it is significant to recognize the potential areas where errors can creep in while designing and executing observational studies.

The most prominent problem in analysis is bias. Bias, is defined as an error in design or execution of a study that influences and consistently distorts the results in one direction. Bias does occur even in randomized controlled trials, but tends to be lesser when compared to observational studies. This overview attempts to analyze the strategies to avoid these biases from the point of sample selection,

design, and final statistical analysis to enhance the value of an observational study.[3,4]

SAMPLING

In an ideal situation, if the subjects are randomly assigned to the treatment or control groups, the effects caused by the intervention is comparable to that of population, thereby ensuring lesser impact of the other interfering factors (co-variables) on the result.[5-7] Thus, in a randomized experiment, groups prior to the intervention should be comparable to ensure that differences in outcomes reflect only the effect of intervention. Random assignment is a way to control extraneous variables and tends to produce groups that are fairly similar on average. In a nonrandomized group, it is often the unmeasured characteristics or extraneous variables that interfere with the interpretation of the results. Precisely, random assignment ensures that the differences between study and control prior to treatment are by chance alone-the flip of a coin in assigning one subject to treatment, another to control-so if a common statistical test rejects the hypothesis that the difference is due to chance, then a treatment effect is demonstrated.[8]

Experiments with harmful and unwanted treatments are to be avoided since they become impractical when subjects refuse to accept the treatment assigned by the researcher. Under such circumstances, the effects of treatments can be examined by an observational study. Differing outcomes may reflect the initial differences rather than effect of the treatment.[9,10] Differences arising out of the factors that influence the treatment decision are called selection biases. The selection bias are of two types: those that can be accurately measured or quantified, called overt biases, and those that cannot be measured but can only be suspected to exist, called hidden biases. Reducing these overt biases and addressing uncertainty related to hidden biases improve the results of observational studies.

Few aspects of an observational study resemble a randomized experiment. Meticulous sampling of the representative subjects from an observational study should improve the results to be generalized to the target population. The researcher should ascertain the status of individuals in different groups based on exposure to the intervention, for example, the exposed (E) and the unexposed (non E) to treatments. In order to assess the effect of the treatment, the population characteristics of the two groups should be comparable. Various measures and calculations are used in observational studies to circumvent them. In studies to estimate the characteristics of a population, in case the samples are not appropriately chosen, the data may not reflect the true value of the population, as the sample may not be representative of the true status of the entire population. On the contrary, few observational studies are applicable universally and can be extrapolated. The risk of a fatal accident on driving after consuming alcohol and the pattern of adverse reactions in patients on Methotrexate (MTX) are some of the examples. In both these conditions, other variables can influence the strength of association. In the first example, road condition, alcoholism and behavior of the population are the influencing factors. With reference to the adverse reaction of MTX, the incidence of hepatitis B/C infection can affect the association. Confounding factors are defined as the characteristics of the population and environment of the observation that can influence the outcome of the study. In an observational study, confounding factors plays a role as there not able to be controlled by investigator. On contrary to the experimental studies, the confounders can be effectively controlled. Therefore, appropriate and a representative sample is necessary for evaluating the treatment outcome. Results of observational studies, from nonrandomized samples, cannot be generalized to the entire population, if the samples are not chosen appropriately to represent the population at large. However, when samples are randomly chosen after defining an appropriate sampling frame, the laws of probability guarantee that the proportions are equal to that of population. The laws of probability also ensure that even in circumstances where the sample data and statistics are not right on target, they will be closer if the sample sizes are large and sample selection follows the representative segments of population. A representative random sample can statistically derive a true population value.

In an intervention-based observational study, researcher may not have the opportunity to randomly allocate exposure and even sometime may have difficulty in picking samples at random from a population. If an appropriate sampling from the target population (framed sample) is available, there is a possibility to either misclassify the exposure of individuals to a risk factor or misdiagnose the disease status or both. If our sampling is not random, it is possible that our methods for estimating the association will give incorrect results, even if the sample has no errors of classification of exposure or disease status. In each of these situations, there is a serious risk of bias in the estimation of the association between exposure and disease.

DESIGN OF OBSERVATIONAL STUDIES

Observational studies are sometimes referred as quasi-experimental or a natural experiments. These terms indicate the perception that even an observational study itself is an experiment. However, one of the distinctive features of

such studies is that the intervention is more naturalistic than planned and controlled. Thus, a well-planned or designed observational study can produce a good result, close to an experiment. An intervention, as defined, can be a treatment, a program, a policy or any other intervention which, in principle, may be applied or withheld from a subject under study. A variable measured prior to treatment, which is not affected by the treatment, is called covariate. A variable measured after treatment that changes by the intervention or treatment is the outcome. An analysis that does not differentiate between covariates and outcomes introduces bias factor to the result, even when none existed. Hence, while planning an observational study, one should attempt to identify all the variables which could influence the outcome.[3,8,10] Some of the variables and the scenarios that can influence the study are discussed in the following sections:

- *Key covariates and outcomes are available for treated and control groups:* The basic component of an observational study is the measurement of important covariates prior to the treatment. Careful data collection should include events occurring over a period of time, as in a longitudinal study, that will define a temporal sequence of events. There should be a distinction between covariates and outcomes. In contrast, data collected in a css is at a single point of time and is often based on interview. In that case, the distinction between covariates and outcomes depends critically on the subjects' recall. Hence, the data may not have the expected sharpness, especially with variables like symptoms of past diseases, experiences, moods, habits, and events. This creates the recall bias, one of the inherent weaknesses of any cross-sectional study. Some of the variables, like age and sex, remain stable whenever they are measured.

- *Haphazard treatment assignment rather than self-selection:* Deliberate selection of subjects can lead to substantial biases in observational studies. For instance, in a study attempting to compare the outcome of use of steroid in rheumatoid arthritis (RA), absence of randomization and rheumatologist's choice to use steroid in patients with more severe RA, can lead to a bias. This pretreatment bias will invariably skew the results since both case and control are not comparable. Hence, the study should reduce this bias by careful choice of the subjects to be included in the study. This can be achieved by using 'special populations offering reduced self-selection', that is, restriction of the study population to predefined subpopulation of the disease may diminish the bias, although it may not completely eliminate bias due to self-selection. In the above example of RA, if only those patients having moderate to mild disease are

selected and the role of steroid use is considered, the bias of self-selection of severe disease to steroid may be reduced. Nevertheless, bias of selection to steroid due to other reasons may still exist.

- *Biases of known direction:* In some settings, the direction of unobserved biases is quite clear even if their magnitude is not. For example, a treatment effect may look like an unambiguous conclusion. An obvious bias may be working in favor of such conclusion. For example, if only patients with severe disease activity of ra are recruited, the changes that occur even with introduction of symptom modifying drug like NSAID may show significant difference than placebo. If this is not considered while assessing the changes, the result may bias toward the arm with the drug.

REDUCING BIASES

Randomization generates treatment and control groups that are comparable prior to treatment. When the sample is not randomized, an understanding of the context becomes important. It is critical to identify the covariates to be considered in the context of the study or the settings. This can create subpopulations with reduced selection biases or determine the direction of hidden biases. As some of the biases may not get eliminated, even the most carefully designed observational study will have weaknesses and ambiguities. A single observational study is often indecisive and replication is often necessary. In replicating an observational study, one should seek to replicate the actual treatment effects, if any, without replicating any biases that may have affected the original study. Some strategies to achieve this are discussed by Cochran.[7]

Generally, three types of bias are distinguished in epidemiological studies: confounding, selection, and information bias. Confounding bias is distinguished from the other two and can be corrected by advanced mathematical methods by estimating association between exposure and disease. The effect of selection and information biases can in no way be removed in the data analysis stage. Thus, one needs to take adequate steps at the initial designing of the study as well as in the execution stages to avoid both selection as well as information bias in an observational study.[3,8]

Confounding bias results when the covariates are mixed up with each other and it is very difficult to isolate or distinguish the effect of each factor individually. For example, in a study to estimate the association of smoking as a risk factor for heart disease, it is possible that people who are exposed to smoking have a higher occurrence of heart disease. However, the study sample may not be as clean as it is expected. It is possible that people who smoke also drink alcohol; so is it

the smoking or alcohol consumption or both responsible for the heart disease? Unless these effects are untangled with advanced statistical methods, derivation appears to be biased by the additional effect of alcohol consumption.

Selection bias results in distortion in the estimate of association between risk factor and disease that result from subject selection. Selection bias could occur because the sampling frame is different from the target population or the sampling procedure has failed to deliver a sample that is a similar to the sampling frame.

Information bias is the distortion in the estimate of risk factors and disease due to systematic measurement error or misclassification of subjects on one or more variables. It is important to realize that these errors are not natural and they are occurring because the physicians or researchers are not careful. These biases do not arise out of a random measurement error or misdiagnosis of an individual, but it is the method of measurement or classification that is the critical issue because it systematically exerts an effect on each of the measurements in the sample.

METHODS FOR REGULATING BIAS

A confounding factor is one that is related to both the exposure and the outcome variables and that does not lie on the causal pathway between them. Ignoring a confounding variable, while assessing the association can lead to either overestimate or underestimate the true association between exposure and outcome, and may even change the direction of the observed effect.

Control of Confounding Variable at the Design Stage

1. Randomization reduces potential for confounding by generating groups that are fairly comparable with respect to known and unknown confounding variables.
2. Restriction eliminates the variable that can interfere as a confounder (for example, recruiting only males into the study will eliminate the confounding variable, sex of the subject).
3. Matching involves selection of a comparable group that is forced to resemble the index group with respect to the distribution of one or more potential confounders.

Among the methods of adjustment for confounding or biases, the most direct and intuitive is matching, which compares each treated individual to one or more controls who appear comparable in terms of observed covariates. Matched sampling is easier when a small group is available together with a large reservoir of potential controls and obtaining data from controls is not prohibitively expensive. In such cases, the standard errors of estimated treatment effects can be substantially reduced by matching each treated subject to several controls (1:4 or more).[11-15]

Control of Confounding Variable at the Analysis Stage

Conventional Approaches

Stratified analysis: In this method, data is divided into subgroups to fix the level of the confounding variable, such that the confounder does not vary within the subgroup and one can evaluate the exposure-disease association within each stratum of confounder. The stratified analysis works best when there are few strata, i.e. only one or two confounders have to be controlled.

Multivariate analysis: If the number of potential confounders is large, multivariate analysis offers the best solution. A multivariate analysis, such as regression analysis, logistic regression, Cox regression, etc. can handle large number of confounders simultaneously.[16]

Newer Approaches to Control Confounders

Graphical approaches: Directed acyclic graph (DAG) is a straightforward tool for relating observed variables and their statistical associations to causal effect. These diagrams are much useful for identifying and controlling the confounders when multivariate analysis fails to yield result. DAGs have vertices representing events and edges representing causal relation between events. DAGs are often acyclic with vertices arranged in the linear order of time, all arrows point in the same direction as time from parent to child (due to causality affecting the future, not the past), and therefore, do not have loops.

Propensity scores: There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatment on outcomes. In observational studies, treatment selection is often influenced by subject characteristics. Consequently, baseline characteristics of treated subjects often differ systematically from those of untreated subjects. Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects while estimating the effect of treatment on outcomes. Historically, researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between the groups. Recently, there has been an increasing interest in methods based on the 'propensity score' to reduce or eliminate the effects of confounding variables in observational data. The propensity score is the probability of treatment assignment conditioned on observed baseline characteristics. The propensity score allows one to design and analyze an

observational (nonrandomized) study to mimic some of the randomized controlled trials. In a nut shell, the propensity score is a balancing score where the distribution of observed baseline covariates will be distributed or balanced out to be similar over treated and untreated subjects. Propensity scores are an alternative method to estimate the effect of treatment when random assignment of treatments to subjects is not feasible. Propensity score matching (PSM) refers to the pairing of treatment and control units with similar values of propensity score, and possibly other covariates, and discarding all unmatched units.[13,17] It is normally used to compare two groups but can be applied to analyze more than two groups. For example, an investigator likes to compare the two treatments (new *vs* standard) in an observational setting in RA with a primary end point of change in the DAS-28-ESR score at 6 months. In a randomized trial, the sample size was estimated to be 448 for 5% level of type I error, 80% of statistical power, but for a study where randomization is less likely to be done, the sample size would increase to 560 for 80% of propensity score, overlapping at 80% statistical power. By using propensity score, the effect of nonrandomization has been compensated by an increase in the number.

Instrumental variables: Instrumental variables (IVs) are used to control confounding and measurement error in observational studies. They allow for the possibility of making causal inferences with observational data. Like propensity scores, IVs can adjust for both observed and unobserved confounding effects.

Marginal structural models: Longitudinal studies in which exposures, confounders, and outcomes are measured repeatedly over time have the potential to allow causal inferences about the effects of exposure on outcome. There is particular interest in estimating the causal effects of medical treatments (or other interventions) in circumstances in which a randomized controlled trial is difficult or impossible. However, standard methods for estimating exposure effects in longitudinal studies are biased in the presence of time-dependent confounders affected by prior treatment. Use of marginal structural models is more efficient to estimate exposure or treatment effects in the presence of time-dependent confounders affected by prior treatment.

PROBLEMS IN ADJUSTMENTS

The term over-adjustment is sometimes used to describe the control (for example, regression adjustment, stratification, restriction) for a variable that increases rather than decreases net bias, or that decreases precision without affecting bias. In many situations adjustment can increase bias. This may be due to reduction in the total causal effect by controlling

for an intermediate variable In the second case, which we term as 'unnecessary adjustment' of an effect estimate, may cause a difference between the uncontrolled and controlled effect estimates, even though no systematic error is present. Moreover, adjusting for surrogates (proxies) of intermediate variables, either ascending or descending, when the desired intermediate variable itself is unmeasured, can have different effects on measures of association depending on the nature of the proxy.

SUMMARY

In the design of an observational study, an attempt should be made to reconstruct some of the structure and strengths of an experiment. Analytical adjustments, such as matching, are used to control for overt biases, that is, pretreatment differences between treated and control groups that are visible in observed covariates. There are many possible sources of error that can result in systematic distortions of study results. These distortions are a problem, especially when the researcher is estimating the association between a risk factor and a health problem. When a risk factor or a protective factor goes undetected, or a behavior or condition is unidentified as a risk or protective factor, the implications can be serious. An erroneously identified risk factor may cause unwanted pain and worry among the public, and perhaps an unnecessary diversion of research funds. Researcher conducting observational studies (cohort, cross-sectional and, especially, case-control) need to be aware of the potential for biases and exert extra caution to eliminate or reduce their effect.

REFERENCES

1. Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990;19(1):205-213.
2. Pearce NE. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993;22(6):1189-1192.
3. Campbell DT. Methodology and epistemology for social science: selected papers. University of Chicago Press, Chicago,1998 p. 315-333.
4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66:688-701.
5. Angrist JD. Randomized trials and quasi-experiments in education research. NBER Reporter, Summer 2003;11-14.
6. Meyer BD. Natural and quasi-experiments in economics. J Business and Economic Statistics 1995;13:151-161.
7. Cochran WG. The planning of observational studies of human populations (with discussion). J Royal Statistical Society Series A 1965;128:134-155.
8. Rosenbaum PR. Observational Studies. 2nd ed. Springer-Verlag, New York NY; 2002.
9. Boruch R. Randomized experiments for planning and evaluation. Sage Publications. Thousand Oaks CA; 1997.

10. Cook TD, Shadish WR. Social experiments: some developments over the past 15 years. Ann Rev Psychol 1994;45:545-580.

11. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics 1968;24(2):295-313.

12. Bergstralh EJ, Kosanke JL, Jacobsen SL. Software for optimal matching in observational studies. Epidemiology 1996;7(3): 331-332.

13. Rubin DB. Using multivariate matched sampling and regression adjustment to control bias in observational studies. J Am Stat Assoc 1979;74:318-328.

14. Smith HL. Matching with multiple controls to estimate treatment effects in observational studies. Socio Methodol 1997;27:325-353.

15. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998;54(3):948-963.

16. Rosenbaum P, Rubin D. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79:516-524.

17. Lu B, Zanutto E, Hornik R, Rosenbaum PR. Matching with doses in an observational study of a media campaign against drug abuse. J Am Stat Assoc 2001;96:1245-1253.